

AI-BASED FILTERING OF VESSEL-INDUCED NOISE IN SIPHON-TYPE RIVER WATER LEVEL MEASUREMENTS AT PHU AN HYDROLOGICAL STATION

Le Van Sang^{1,*}, Le Van Phan², Le Do Van Bang³

¹University of Phan Thiet, Phan Thiet City, Vietnam

²Nong Lam University, Ho Chi Minh City, Vietnam

³Assumption University, Bangkok City, Thailand

*Email: lhsang@upt.edu.vn

Received: 19 January 2026; Revised: 9 March 2026; Accepted: 17 April 2026

ABSTRACT

Accurate river water level observations are fundamental for hydrological analysis, tidal assessment, and flood forecasting. At the Phu An Hydrological Station, water level measurements recorded by a siphon-type gauge system are frequently contaminated by high-frequency oscillations caused by vessel traffic on the river. These vessel-induced disturbances introduce significant noise into the observed time series, leading to errors in identifying tidal extrema and reducing the reliability of water level data, particularly under increasing hydrodynamic variability associated with climate change. An artificial intelligence (AI)-based filtering framework is proposed to mitigate vessel-induced noise in river water level observations. The proposed approach integrates signal preprocessing and deep learning techniques to model the intrinsic temporal dynamics of natural river water levels while suppressing non-hydrological fluctuations. Specifically, a deep learning model is trained to learn the smooth and continuous behavior of water level variations driven by tides and river flow, enabling the reconstruction of a standardized and noise-free water level hydrograph from raw observations affected by vessel-induced disturbances. The experimental results demonstrate that the proposed AI-based method effectively reduces high-frequency noise and significantly improves data stability without distorting the underlying hydrological signal. The corrected water level series shows improved accuracy in capturing tidal peaks and troughs, thereby improving the reliability of hydrological observations at Phu An Station. The proposed methodology provides a practical and scalable solution for improving water level measurements at river stations influenced by intensive vessel traffic and offers strong support for hydrological forecasting, climate change impact assessment, and sustainable water resource management.

Keywords: River water level monitoring, Water level noise filtering, Vessel-induced disturbance, Time-series analysis, Tidal water level, Hydrological data quality.

1. INTRODUCTION

River water level monitoring plays a particularly important role in water resource management, disaster prevention, and the operation of irrigation works, especially in large cities with high population density and infrastructure. In practice, many water level monitoring stations use siphon wells to limit the direct impact of flow and environmental influences on the measuring equipment, thereby improving the stability of monitoring data.

In the downstream area of the Saigon-Dong Nai river system, the Phu An Hydrological Station (Ho Chi Minh City) is one of the important water level monitoring stations, providing

data for tidal monitoring, flood warning, and water resource management for the central and surrounding areas of Ho Chi Minh City. However, due to its location on a river with high water traffic density [1], [2], water level data measured at the Phu An Station is frequently affected by waves from ship activity, causing short-term interference in the measurement signals.

Noise from vessels at Phu An Station is characterized by short durations, small amplitudes, but high frequency, causing the water level signal to exhibit sawtooth-like oscillations. If not properly addressed, this noise can lead to inaccuracies in determining the actual water level, directly affecting the reliability of observational data and thus reducing the effectiveness of tidal surge and flood warning systems in Ho Chi Minh City.

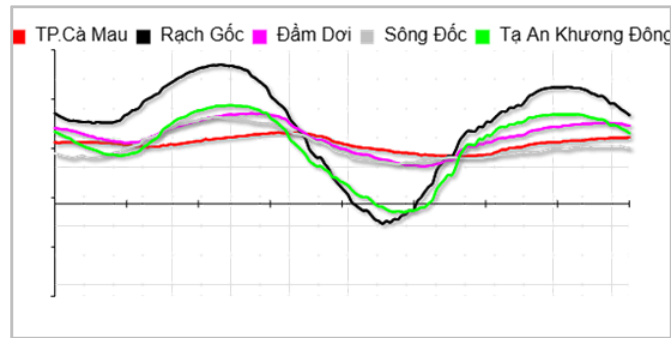


Fig. 1. Current status of water level issues at hydrological stations without noise processing

In the past, many signal noise processing methods have been applied to hydrological data, such as moving average filtering, low-pass filtering, or classical digital filtering methods. However, for water level data at Phu An Station, which has a high sampling period and is heavily influenced by field noise, single-stage filtering methods often have limitations in balancing noise removal while preserving the slow trend of water level change. Signal delays or distortion of water level trends can affect data exploitation in warning and forecasting applications [3], [4], [8].

Based on observations at the Phu An Hydrological Station, this paper proposes a two-stage noise reduction method for river water level data from siphon wells. The method combines a median filter to eliminate short-term noise from vessels and a Kalman filter to accurately estimate the water level trend over time. The method was validated using real-world data from several days at the Phu An Station with a 5-minute sampling interval, demonstrating significant effectiveness in reducing noise and improving the quality of water level data for monitoring and warning purposes in the Ho Chi Minh City area.

2. OVERVIEW OF NOISE REMOVAL METHODS IN WATER LEVEL MEASUREMENT DATA

River water level measurement data is often affected by various noise sources such as flow oscillations, wind effects, ship activity, and measurement equipment errors. Among these, noise from ships is characterized by its sudden appearance, short duration, and high frequency, causing sawtooth-like oscillations in the measured water level signal. To handle this type of noise, many signal filtering methods have been studied and applied in practice.

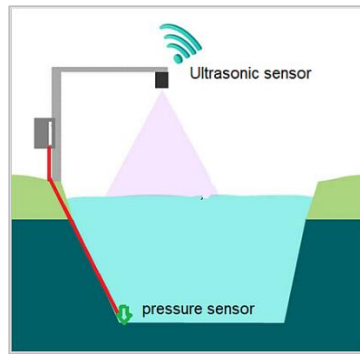


Fig. 2. In-situ hydrological monitoring station.

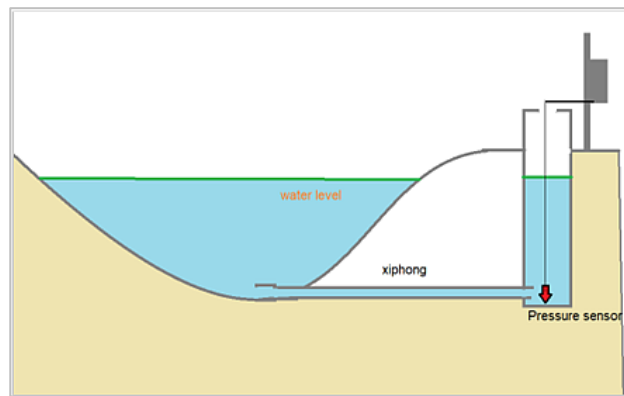


Fig. 3. Siphon-based hydrological measurement station.

2.1. Moving Average Filter

The Moving Average Filter is a simple and common method in which the signal value at a given time is replaced by the average value of several neighboring samples. This method has the advantage of being easy to implement and effective in reducing random noise. However, when applied to water level data with a high sampling period, moving average filtering often causes signal lag and degrades important characteristics of the water level trend.

2.2. Median Filter

The median filter is a nonlinear filtering technique in which the signal value at a given time is replaced by the median value of a neighboring sampling window. This method is particularly effective in removing impulse noise or short-term oscillations with unusual amplitudes. For river water level data affected by boat waves, median filtering helps eliminate sawtooth noise peaks while preserving the slow trend of water level change.

However, although median filtering effectively removes short-term noise, it does not explicitly estimate the underlying system state nor provide a mechanism for optimal state estimation over time

2.3. Kalman Filter

Kalman filtering is a method for estimating the optimal state based on state-space models and probability theory. The Kalman filter allows the combination of information from the system model and measurement data to estimate the actual state of the system over time. This method has been widely applied in signal processing and control problems, particularly effective for systems with slowly changing dynamics.

In the problem of river water level measurement, the Kalman filter can estimate the actual water level trend and smooth the measurement signal. However, when the input signal contains a lot of short-term noise with large amplitudes, the effectiveness of the Kalman filter may be reduced. Recent studies have shown that Kalman-based approaches remain effective for hydrological time-series smoothing when appropriately combined with preprocessing techniques [5], [6].

2.4. Remarks

From the above analysis, it can be seen that individual filtering methods all have certain limitations when applied to river water level measurement data with high sampling periods and strong influence from noise from ships. Therefore, combining filtering methods to take advantage of the strengths of each method is necessary to improve the quality of measurement data.

3. PROPOSED NOISE PROCESSING METHOD

3.1. Long Short-Term Memory (LSTM) for Water Level Noise Filtering

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) architecture specifically designed to model temporal dependencies in sequential data. Unlike conventional feedforward neural networks, LSTM incorporates memory cells and gating mechanisms that enable the network to retain long-term information while selectively forgetting irrelevant short-term fluctuations. This characteristic makes LSTM well suited for time-series analysis tasks involving nonlinear and complex temporal patterns [4], [6].

In the context of river water level observations, LSTM has been increasingly explored to model water level dynamics driven by tidal effects and river flow variations. By learning temporal correlations from historical data, an LSTM model can approximate the underlying smooth and continuous behavior of water level changes, thereby reducing the influence of high-frequency noise. When applied to noisy water level measurements, the LSTM model implicitly acts as a data-driven filtering mechanism that reconstructs a smoothed water level series based on learned temporal patterns [7].

However, despite its strong modeling capability, the application of LSTM for noise filtering in operational hydrological monitoring systems presents several challenges. First, LSTM models require a large amount of representative training data to achieve stable performance, and the quality of the output strongly depends on the selection of training periods and network hyperparameters. Second, the computational complexity of LSTM-based approaches is significantly higher than that of traditional filtering methods, which may limit their applicability in real-time processing environments at hydrological stations. In addition, the black-box nature of deep learning models reduces interpretability, making it difficult to directly relate the filtering behavior to physical hydrological processes [9].

In this study, LSTM is considered as a reference data-driven approach for comparison purposes rather than as the primary noise processing solution. The proposed median-Kalman filtering framework is designed to provide a computationally efficient, interpretable, and practically deployable alternative for suppressing vessel-induced noise in siphon-type river water level measurements. The comparison with LSTM-based results helps highlight the advantages of the proposed method in terms of robustness, real-time applicability, and preservation of hydrological trends under real operational conditions.

3.2. Water Level Measurement Signal Model

The water level signal measured from the siphon well system can be modeled as follows:

$$y_k = x_k + v_k, \tag{1}$$

where y_k is the measured water level at time k , x_k is the actual water level to be estimated, and v_k is the measurement noise, mainly generated by the impact of ships and environmental factors.

The goal of the problem is to accurately estimate x_k from the measurement signal sequence y_k .

3.3. Median Filtering to Remove Short-Term Noise

In the first processing stage, a median filter is applied to remove short-term noise from ships. With a filter window size of $2M + 1$, the median signal after filtering is determined by:

$$\tilde{y}_k = \text{median}(y_{k-M}, y_{k-M+1}, \dots, y_{k+M}). \tag{2}$$

In this study, the window size was chosen as 9 samples, corresponding to a time interval of 45 seconds, consistent with the characteristics of waves caused by ships in the actual measurement data.

3.4. Kalman Filter for Estimating Water Level Trend

After removing short-term noise, the signal \tilde{y}_k is fed into a Kalman filter to estimate the actual water level trend. The state model is chosen as follows:

$$x_k = x_{k-1} + v_k \tag{3}$$

$$\tilde{y}_k = x_k + v_k \tag{4}$$

where v_k is the system noise reflecting the slow change of water level over time, and v_k is the remaining measurement noise after median filtering. These two noises are assumed to be white Gaussian noise with variances Q and R , respectively.

The parameters Q and R are selected based on the characteristics of actual measured data and are fine-tuned experimentally to achieve optimal filtering efficiency.

3.5. Proposed Method Block Diagram

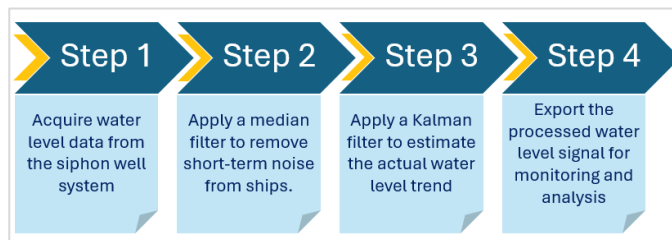


Fig. 4. Proposed Method Block Diagram

The proposed noise processing method is designed as a two-stage filtering framework for river water level measurements obtained from a siphon-type gauge system. The overall processing flow, as illustrated in Figure 4, aims to suppress vessel-induced short-term disturbances while preserving the intrinsic temporal behavior of the river water level.

The input of the proposed method is a time-series of river water level observations collected from the siphon well system. These measurements are sampled at a high temporal resolution and contain both the actual hydrological signal and high-frequency noise generated by vessel traffic and environmental effects. Prior to advanced estimation, the raw data are

directly processed through the proposed filtering framework without the need for hardware modification or complex preprocessing.

The proposed method consists of the following main processing stages:

- Median filtering stage:

In the first stage, a median filter is applied to the raw water level signal to eliminate short-term disturbances caused by ship-induced waves. The median filter effectively suppresses impulsive and high-frequency noise while maintaining the continuity of the water level signal. This stage serves as a preprocessing step that improves the robustness of subsequent estimation by reducing abnormal fluctuations.

- Kalman filtering stage:

In the second stage, the median-filtered signal is fed into a Kalman filter to estimate the actual water level trend over time. The Kalman filter models the water level as a slowly varying state and combines system dynamics with measurement information to produce an optimal estimate. This stage enhances signal smoothness and accurately captures long-term tidal and river flow variations.

- Output signal generation:

The output of the proposed framework is a noise-reduced and smoothed water level time series that preserves the essential hydrological characteristics. The processed signal is suitable for real-time monitoring, tidal analysis, flood warning, and further hydrological applications.

By integrating median filtering and Kalman filtering in a two-stage structure, the proposed method effectively exploits the complementary strengths of both techniques. This approach significantly improves noise suppression performance under real-world measurement conditions while maintaining computational efficiency and practical applicability for operational hydrological stations [10].

4. EXPERIMENTS AND EVALUATION

4.1. Description of Water Level Dataset

The data used in this study are river water level measurements collected from a siphon well system over several consecutive days, with a 5-minute sampling interval. Water levels are stored as time series data in centimeters (cm). The total number of data samples collected exceeds 400,000, fully reflecting the changing water levels and the impact of noise from boat activity on the river.

Before processing, the data were normalized and sorted in ascending chronological order. Duplicate or invalid samples (if any) were removed to ensure the continuity and accuracy of the data series.

Due to the large data volume, during the algorithm evaluation process, the paper selected several representative data segments ranging from a few hours to a few days, including periods with different levels of noise: strong noise, medium noise, and low noise. This selection method helps to comprehensively evaluate the effectiveness of the proposed method under different real-world measurement conditions.

4.2. Experimental Configuration and Benchmark Methods

To evaluate the effectiveness of the proposed noise reduction method, the following methods were implemented and compared:

- P1: Original water level signal (unprocessed)
- P2: Moving average filter

P3: LSTM model

P4: Median filter

P5: Kalman filter

P6: Proposed method (Median filter + Kalman filter)

For the median filter, the window size was selected as 9 samples, corresponding to a time interval of 45 seconds. For Kalman filtering, the system noise variance parameters Q and the measured noise variance R are selected and fine-tuned experimentally to achieve the best signal smoothing effect without losing the trend of water level change.

The methods are evaluated based on both quantitative and qualitative criteria.

4.3. Evaluation Criteria

The criteria used to evaluate the effectiveness of noise removal include:

- Standard deviation of the signal: reflects the degree of fluctuation of the water level after processing.
- Ability to preserve the water level trend: evaluated by observing the fit of the filtered signal to the trend of water level change over long periods.
- Signal lag: reflects the degree of lag of the processed signal compared to the original signal.

Since there is no absolute “standard” water level value at all times, the evaluation is mainly based on relative comparisons between methods and visual analysis combined with statistical indicators.

4.4. Results and Discussion

Experimental results showed that the original water level signal exhibited many short-term sawtooth-like oscillations, particularly noticeable during periods when ships were moving through the measurement area. Moving average filtering helped reduce some of the noise but caused signal delay and degraded the trend characteristics of the water level.

Median filtering provided better short-term noise removal, especially effective for unusual noise peaks. However, the signal after median filtering still showed slight oscillations and did not clearly show the water level trend over long periods.

Kalman filtering provided better signal smoothing and water level trend estimation compared to the single methods. However, when the input signal still contained a lot of short-term noise, the effectiveness of Kalman filtering could be affected.

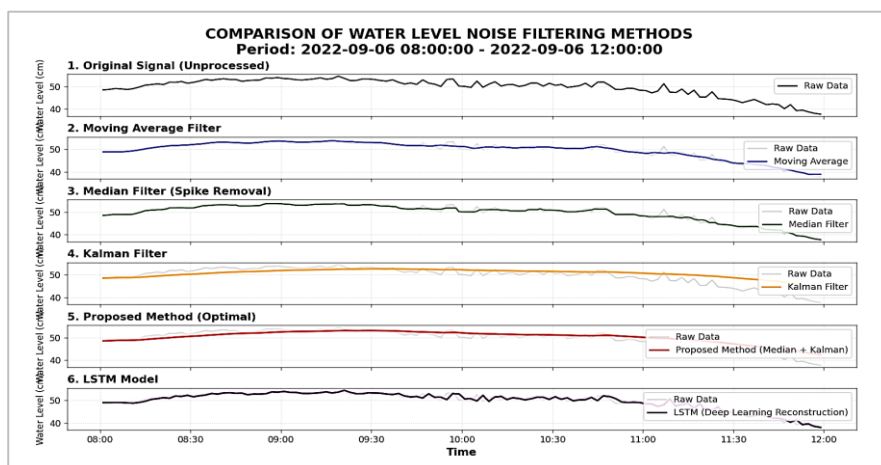


Fig. 5. Comparison of water level noise filtering methods

Over short time intervals, this graph shows a higher magnitude of observed water-level fluctuations.

Table 1. Comparison of river water level measurements at Phu An hydrological station using different filtering methods

Time	Raw Water Level	MA Filter	Median Filter	Kalman Filter	Proposed Method	LSTM AI
06/09/2022 8:01	48.52	48.81	48.52	48.52	48.52	49.04
06/09/2022 8:02	48.61	48.81	48.61	48.54	48.56	49.04
06/09/2022 8:03	48.69	48.81	48.69	48.57	48.59	49.04
06/09/2022 8:04	48.9	48.81	48.9	48.63	48.66	49.04
06/09/2022 8:05	49.1	48.81	48.9	48.7	48.71	49.04
06/09/2022 8:06	49.02	48.81	48.9	48.74	48.75	49.04
06/09/2022 8:07	48.94	48.86	48.9	48.76	48.77	49.04
06/09/2022 8:08	48.79	48.94	48.94	48.76	48.78	48.97
06/09/2022 8:09	48.65	49.05	48.94	48.75	48.79	48.91
06/09/2022 8:10	48.85	49.18	48.94	48.76	48.8	48.84
06/09/2022 8:11	49.06	49.32	49.06	48.78	48.82	48.75
06/09/2022 8:12	49.41	49.49	49.41	48.83	48.87	48.92
06/09/2022 8:13	49.76	49.68	49.76	48.89	48.95	49.09
06/09/2022 8:14	50.17	49.9	50.17	48.98	49.04	49.34
06/09/2022 8:15	50.59	50.16	50.59	49.08	49.16	49.63
06/09/2022 8:16	50.71	50.37	50.71	49.17	49.26	50.02
06/09/2022 8:17	50.83	50.55	50.83	49.26	49.37	50.41
06/09/2022 8:18	51	50.74	50.83	49.36	49.48	50.57
06/09/2022 8:19	51.16	50.95	51	49.45	49.57	50.73
06/09/2022 8:20	51	51.12	51	49.52	49.66	50.96

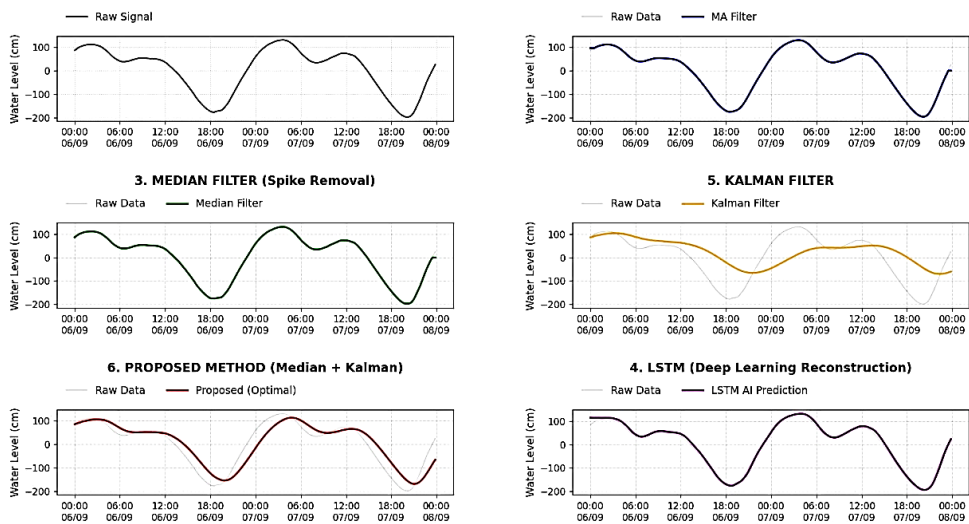


Fig. 6. Water level analysis results at Phu An Station. Period: 06/09/2022 - 07/09/2022

Although short-term oscillations are not clearly observed at the daily resolution, the noise-filtering methods continue to perform effectively.

Illustrates a comparison of different noise filtering methods applied to river water level observations recorded at the Phu An Hydrological Station over several consecutive days. The raw water level signal exhibits pronounced high-frequency oscillations caused by vessel-induced disturbances, particularly noticeable during rapid rising and falling phases of the tidal cycle.

The moving average filter reduces part of the high-frequency noise but introduces noticeable signal delay and attenuates tidal extrema. The median filter effectively removes short-term spike noise; however, the resulting signal still contains minor irregularities and does not fully capture the smooth long-term trend of water level variation. The Kalman filter provides improved smoothing and trend estimation but remains sensitive to short-term disturbances when applied directly to the raw signal.

In contrast, the proposed two-stage filtering method combining median filtering and Kalman filtering produces the most stable and consistent water level signal. The method effectively suppresses vessel-induced high-frequency noise while preserving the natural tidal pattern and accurately capturing peak and trough water levels. These results demonstrate the suitability of the proposed approach for processing high-frequency river water level data affected by intensive vessel traffic.

The proposed method combining median filtering and Kalman filtering yielded superior results compared to the other methods. The processed signal effectively eliminates short-term oscillations caused by vessels, preserves the trend of water level changes over time, and significantly reduces signal delay. This result demonstrates the suitability of the proposed method for river water level data with high sampling cycles and strong influence from field noise.

5. CONCLUSION AND FUTURE DEVELOPMENT

This study proposes and validates a two-stage noise reduction method for river water level data from siphon wells, combining median filtering and Kalman filtering to eliminate short-term noise from vessels and accurately estimate the real-time water level trend. The method was validated on real-world data measured over several days with a 5-minute sampling interval.

Experimental results show that the proposed method significantly improves water level signal quality, effectively reduces noise while preserving the water level trend and minimizing signal delay. The method can be directly implemented in existing river water level monitoring systems without requiring hardware modifications.

In further studies, the method could be extended in areas such as automatically optimizing Kalman filter parameters, incorporating advanced signal analysis methods, or integrating it into real-time water level warning systems.

Acknowledgement: This research was supported by the project entitled “Development of a Digital Technology- and Citizen Science-Based Solution for Extreme Weather Warning under Climate Change Conditions in Ho Chi Minh City.” The project provided supporting data infrastructure, hydrological monitoring resources, and a research platform related to real-time environmental monitoring and warning systems in Ho Chi Minh City.

REFERENCES

- [1] C. R. Haghghi and I. N. A. Haghghi, "Ship-generated wave-induced sediment dynamics in upper St. Lawrence River: Analysis of field data," *Natural Hazards*, vol. 121, no. 14, pp. 16689–16718, 2025, doi: <https://doi.org/10.1007/s11069-025-07445-3>.
- [2] "World Meteorological Organization," Guide to Hydrological Practices, 2021, 2021.
- [3] D. Wang *et al.*, "Sample entropy-based adaptive wavelet de-noising approach for meteorologic and hydrologic time series," *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 14, pp. 8726–8740, Jul. 2014, doi: <https://doi.org/10.1002/2014JD021869>.
- [4] H. Lin, Q. Zhang, S. Band, and H. Ting, "Time series-based groundwater level forecasting using gated recurrent unit deep neural networks," *Engineering Applications of Computational Fluid Mechanics*, vol. 16, no. 1, pp. 1655–1672, Dec. 2022, doi: <https://doi.org/10.1080/19942060.2022.2104928>.
- [5] C. K. Chui and G. Chen, "Correlated system and measurement noise processes," in *Kalman Filtering: With Real-Time Applications*, 5th ed. Cham, Switzerland: Springer, 2017, ch. 4, pp. 53–67, doi: https://doi.org/10.1007/978-3-662-02508-6_4.
- [6] J. M. Frame *et al.*, "Deep learning rainfall–runoff predictions of extreme events," *Hydrology and Earth System Sciences*, vol. 26, no. 13, pp. 3377–3392, Jul. 2022, doi: <https://doi.org/10.5194/hess-26-3377-2022>
- [7] B. B. Sahoo, R. Jha, A. Singh, and D. Kumar, "Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting," *Acta Geophysica*, vol. 67, no. 4, pp. 1471–1481, Oct. 2019, doi: <https://doi.org/10.1007/s11600-019-00330-1>.
- [8] M. Sit, B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir, "A comprehensive review of deep learning applications in hydrology and water resources," *Water Science and Technology*, vol. 82, no. 12, pp. 2635–2670, 2020, doi: <https://doi.org/10.2166/wst.2020.369>.
- [9] S. Mousavi, "Conjugation of deep learning and de-noising data methods for predicting river water levels," *Journal of Hydro-environment Research*, vol. 52, pp. 26-37, Jan 2024, doi: <https://doi.org/10.1016/j.jher.2023.12.002>
- [10] D. Simon, "Kalman filtering with modern applications," *IEEE Control Systems Magazine*, p. 34–53, 2020.